حریری، نجلا؛ طاهری، سیدمهدی (۱۳۹۴). بررسی تأثیر حذف خط اعلان زبان نشانه‌گذاری گسترش‌پذیر (XML) و تغییر پسوند فایل بر افزایش نمایه‌پذیری و پیدانمایی پیشینه‌های فراداده‌ای در محیط موتورهای کاوش وب. *پژوهش‌نامه کتابداری و اطلاع‌رسانی*، ۵ (۱)، ۲۷-۱۰.

بررسی تأثیر حذف خط اعلان زبان نشانه‌گذاری گسترش‌پذیر (XML) و تغییر پسوند فایل بر افزایش نمایه‌پذیری و پیدانمایی پیشینه‌های فراداده‌ای در محیط موتورهای کاوش وب

دکتر نجلا حریری [1]، دکتر سید مهدی طاهری [2]

تاریخ دریافت: ۱۳۹۱/۷/۱۷ تاریخ پذیرش: ۱۳۹۱/۱۰/۱۲

چکیده

هدف: *هدف از این پژوهش تبیین تأثیر حذف خط اعلان زبان نشانه‌گذاری گسترش‌پذیر و تغییر پسوند فایل بر افزایش نمایه‌پذیری و پیدانمایی نام‌های برچسب عناصر فراداده‌ای هسته دوبلین، مارک ۲۱، و طرح فراداده‌ای توصیف شیء در محیط موتورهای کاوش وب بود.*

روش: *پژوهش با روش تجربی انجام شد. جامعه پژوهش را شش‌صد پیشینه فراداده‌ای در قالب دو گروه آزمون و گواه تشکیل می‌دادند. گروه گواه دارای سیصد پیشینه، تولیدشده بر مبنای فرانمای استاندارد طرح‌های فراداده‌ای موردمطالعه، و گروه آزمون نیز دارای سیصد پیشینه، ایجادشده بر مبنای فرانماهای یادشده اما دارای ویژگی‌های خاص، بدون سطر اعلان ایکس‌ام‌ال و با پسوند فایل همخوان با نام طرح فراداده‌ای مرتبط بودند.*

یافته‌ها: *نتایج پژوهش نشان داد که موتورهای کاوش گوگل و یاهو نام‌های برچسب تمامی عناصر فراداده‌ای مربوط به پیشینه‌های عضو گروه آزمون را نمایه‌سازی نموده، و در نتایج جستجوی خود پدیدار ساختند. اما چنین واکنشی را به نام‌های برچسب عناصر پیشینه‌های فراداده‌ای گروه گواه نشان ندادند. نمایه‌سازی نام‌های برچسب عناصر پیشینه‌های فراداده‌ای گروه آزمون، و پدیدار شدن آنها در نتایج جستجوی موتورهای کاوش، امکان بازیابی پیشینه‌های فراداده‌ای یادشده از طریق نام‌های برچسب آنها را فراهم نموده است. چنین امکانی برای پیشینه‌های گروه گواه میسر نیست. در پایان مقاله، الگوهایی ارائه شده است.*

کلیدواژه‌ها: *پسوند فایل، پیدانمایی، خط اعلان ایکس‌ام‌ال، طرح فراداده‌ای توصیف شیء، طرح فراداده‌ای هسته دوبلین، قالب فراداده‌ای مارک ۲۱، موتورهای کاوش، نام‌های برچسب عناصر، نمایه‌سازی*

---

۱. دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران

۲. دانش آموخته مقطع دکترای علوم کتابداری و اطلاع رسانی دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران

مقدمه

فراداده ابزاری کارآمد برای توصیف و سازماندهی اشیای محتوایی شبکه‌ای است. این ابزار برای بازنمون خود نیاز به بستر نحوی دارد. بستر نحوی عبارت است از مجموعه‌ای از قواعد، دستورالعمل‌ها، و نشانه‌ها برای اعمال ساختاری خاص بر روی محتوای متنی به‌منظور ذخیره‌سازی، فهم و انجام پردازش‌های خاص توسط ماشین (رایانه). انتخاب بستر نحوی برای پیاده‌سازی پیشینه‌های فراداده‌ای بر روی کارکردها، و به‌ویژه تعامل آنها با دیگر نظام‌ها ازجمله موتورهای کاوش و حتی دیگر طرح‌های فراداده‌ای بسیار تأثیرگذار است.

تمایل به استفاده از زبان نشانه‌گذاری گسترش‌پذیر (XML) به‌عنوان بستر نحوی از سوی طراحان و پیاده سازان استانداردهای فراداده‌ای در راستای افزایش کارکردپذیری، و به‌طور خاص، تعامل آنها با دیگر نظام‌های اطلاعاتی بوده است (Qin, 2000; Luk et al., 2002 Bogaards, 2004; Gigee, 2006; Farance and Gillman, 2007; Tabatabaie Amiri, 2012; Taheri and Hariri, 2012). تغییر بستر نحوی قالب فراداده‌ای مارک ۲۱ به زبان نشانه‌گذاری گسترش‌پذیر، و توسعه مشتقات آن از جمله طرح فراداده‌ای توصیف شیء (MODS) بر پایه بستر یادشده، و نیز انتخاب قالب ایکس‌ام‌ال به‌عنوان یکی از قالب‌های اصلی بازنمون طرح فراداده‌ای هسته دوبلین، به‌عنوان مهم‌ترین و پرکاربردترین استانداردهای فراداده‌ای در همین راستا بوده است.

اشیای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر، اشیای متنی با نشانه‌گذاری اضافی قلمداد می‌شوند (Bray et al., 2008). ازسوی‌دیگر، درحال‌حاضر سه رویکرد کلی برای بازیابی اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر در نظام‌های بازیابی اطلاعات وجود دارد. رویکرد مبتنی بر پایگاه داده‌ها، رویکرد مبتنی بر بازیابی اطلاعات ، و رویکرد دووجهی (ترکیبی از دو رویکرد قبلی) (Luk et al., 2000; Luk et al., 2002). بر پایه رویکرد مبتنی بر بازیابی اطلاعات، برای پرداختن به برچسب‌های نشانه‌گذاری چندین روش وجود دارد. یک روش، حذف همه برچسب‌هاست. مزیت این روش، سادگی آن است، و زیان آن از دست دادن داده‌ها، و کاهش کارآیی بازیابی است. این رویکرد، رویکرد کنونی موتورهای کاوش است. روش دیگر، استخراج داده‌های ساختاری و بافتی از اشیای محتوایی (داده و فراداده) مبتنی بر ایکس‌ام‌ال برای نمایه‌سازی است. و روش جامع‌تر، نمایه‌سازی نام‌های برچسب‌ها به‌عنوان اصطلاحات نمایه‌ای است (Luk et al., 2000; Taheri and Hariri, 2012; Hariri et al., 2014).

در روش اخیر، امکان جستجوهای مبتنی بر عنصر در محیط موتورهای کاوش به وجود می‌آید. به‌عبارت‌دیگر، کاربران می‌توانند دامنه جستجوهای خود را به عناصری خاص محدود نمایند. این قابلیت‌ها

که ناشی از افزایش سطح پردازشی اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر است، بهبود تعامل نظام‌هایی که از این فناوری استفاده می‌کنند با دیگر نظام‌های اطلاعاتی را باعث گردیده، و ایده میان‌کنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب، به‌عنوان پرطرفدارترین ابزارهای دسترسی به محتوای وب را بیش از گذشته مطرح نموده است. این نوع از میان‌کنش‌پذیری به‌صورت معناشناختی در سطح مخازن اطلاعاتی انجام می‌شود (Hirwade, 2011).

در این راستا تلاش‌های ویژه‌ای انجام شده است. نتایج پژوهش‌های مختلف بیانگر بخشی از این تلاش‌هاست. ازجمله، لوک و همکاران (Luk et al., 2000)، طاهری و همکاران (2009)، آقاعابدی (2010)، و طباطبایی امیری و همکاران (2011) در پژوهش‌های خود نشان دادند، انتخاب «زبان نشانه‌گذاری گسترش‌پذیر» به‌عنوان بستر نحوی پیشینه‌های فراداده‌ای، موجب افزایش نمایه‌پذیری و قابلیت بازیابی آنها در موتورهای کاوش می‌گردد. موتورهای کاوش بر پایه رویکرد حذف برچسب اشیای محتوایی مبتنی بر ایکس‌ام‌ال را نمایه‌سازی می‌کنند.

مسئله مهمی که در فرایند میان‌کنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب مطرح می‌شود، چگونگی تغییر رویکرد موتورهای کاوش در نمایه‌سازی و پیدانمایی نام‌های برچسب عناصر فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر است. از یک‌سو، رویکرد کلی موتورهای کاوش در نمایه‌سازی اشیای محتوایی مبتنی بر ایکس‌ام‌ال، حذف برچسب است، و از سوی دیگر، شواهد (ازجمله نسخه «حافظه پنهان (cache)» موتورهای کاوش) نشان می‌دهند، نرم‌افزارهای نمایه‌سازی آنها همه نویسه‌ها و اجزای اشیای محتوایی را نمایه می‌کنند (Search Engine Watch, 2007). بر این اساس، آیا اعمال تغییراتی خاص بر روی بستر نحوی پیشینه‌های فراداده‌ای، با حفظ قابلیت‌های زبان نشانه‌گذاری گسترش‌پذیر، بر نمایه‌پذیری و پیدانمایی نام‌های برچسب عناصر آنها خواهد افزود؟ آیا می‌توان پیشینه‌های فراداده‌ای را بر پایه جستجوهای مبتنی بر عنصر در محیط موتورهای کاوش وب بازیابی نمود؟ پژوهش حاضر به‌منظور پاسخگویی به این مسئله طرح‌ریزی شده است. در همین راستا، و به‌منظور پاسخ به این مسئله، بررسی این فرضیه مدنظر است: حذف خط اعلان شیء مبتنی برای ایکس‌ام‌ال و تغییر پسوند فایل مربوط به پیشینه‌های فراداده‌ای، امکان نمایه‌سازی و پیدانمایی نام‌های برچسب عناصر فراداده‌ای توسط موتورهای کاوش را فراهم می‌سازد.

مرور تحلیلی پیشینه‌ها

تغییر بستر نحوی پیشینه‌های فراداده‌ای توسط طراحان استانداردهای فراداده‌ای از زبان نشانه‌گذاری فرامتن (HTML) به زبان نشانه‌گذاری گسترش‌پذیر، با هدف رهایی از محدودیت‌های بستر پیشین و

استفاده از قابلیت‌های بستر جدید بود. این اقدام استراتژیک به دلیل ویژگی‌های منحصربه‌فرد قالب ایکس‌ام‌ال، میانکنش‌پذیری نظام‌های فراداده‌ای با دیگر نظام‌های اطلاعاتی ازجمله موتورهای کاوش وب را بهبود می‌بخشید. یافته‌های پژوهش‌های پیشینی که در حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب انجام شد، به‌روشنی بیانگر ضعف یا عدم کارایی زبان نشانه‌گذاری فرامتن، و تأثیر مثبت زبان نشانه‌گذاری گسترش‌پذیر بر فرایند میانکنش‌پذیری یاد شده بودند. یافته‌های پژوهش‌هایی که قالب اچ‌تی‌ام‌ال را به‌عنوان بستر نحوی پیاده‌سازی عناصر یا پیشینه‌های فراداده‌ای مورد مطالعه قرار داده بودند، حاکی از آن بود که تنها ارزش‌های فرابرچسب‌های قالب اچ‌تی‌ام‌ال برای موتورهای کاوش وب ارزش نمایه‌سازی دارند، و این موتورها به ارزش‌های عناصر پیشینه‌های تولیدشده بر مبنای طرح‌های فراداده‌ای استاندارد واکنش مثبت نشان نمی‌دهند.

نتایج پژوهش‌های ترنر و برکبیل (Turner and Brackbill, 1998)، کوئودو-توررو (Quevedo-Torrero, 2004)، و ژانگ و دیمیتروف (Zhang and Dimitroff, 2005a) نمایه‌پذیری ارزش‌های برچسب و فرابرچسب‌های زبان نشانه‌گذاری فرامتن را تأیید می‌کنند، و پژوهش‌های سوکویتن (Sokvitne, 2000) و صفری (Safari, 2005) که تأثیر عناصر فراداده‌ای هسته دوبلین در قالب اچ‌تی‌ام‌ال به‌منزله یک طرح فراداده‌ای استاندارد را بر قابلیت بازیابی و بهبود رتبه اشیای محتوایی در میان نتایج موتورهای کاوش بررسی نمودند، کاهش یا عدم نمایه‌پذیری ارزش‌های آن عناصر، و درنتیجه کاهش تأثیر آنها بر بهبود دسترسی و رتبه‌بندی اشیای محتوایی را نشان دادند. برخی دیگر از پژوهش‌هایی که به مطالعه واکنش موتورهای کاوش به ارزش‌های عناصر مبتنی بر زبان نشانه‌گذاری فرامتن اختصاص داشتند، رویکردی ترکیبی و مقایسه‌ای را به برچسب‌ها و فرابرچسب‌های زبان نشانه‌گذاری فرامتن، و عناصر پیشینه‌های طرح‌های فراداده‌ای استاندارد برگزیدند. دلیل انتخاب این رویکرد، بررسی بیشتر تأثیر مثبت فرابرچسب‌های اچ‌تی‌ام‌ال، و تأثیر خنثی عناصر فرداده‌ای مبتنی بر طرح‌های استاندارد بود. این پژوهش‌ها نیز به نتایجی مشابه، یعنی کارایی ارزش‌های فرابرچسب‌های اچ‌تی‌ام‌ال، و عدم یا کاهش کارایی ارزش‌های عناصر مبتنی بر طرح‌های استاندارد در بهبود دسترسی به اشیای محتوایی دست یافتند (Henshaw and Valauskas, 2001; Zhang and Dimitrof 2004; Zhang and Dimitroff, 2005b; Mohamed, 2006; Sharif, 2007).

گروه دیگر پژوهش‌های حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب، زبان نشانه‌گذاری گسترش‌پذیر را به‌عنوان بستر نحوی پیشینه‌های فراداده‌ای مورد مطالعه برگزیدند. یافته‌های پژوهش‌های طاهری و همکاران (2009)، آقاعابدی (2010)، و طباطبایی امیری و همکاران (2011) بر

بهبود نمایه‌پذیری ارزش‌های عناصر مبتنی بر قالب ایکس‌ام‌ال در محیط موتورهای کاوش وب تأکید می‌نمود. اما بستر نحوی ایکس‌ام‌ال بر رتبه‌بندی پیشینه‌های فراداده‌ای در نتایج موتورهای کاوش بی‌تأثیر است. همچنین این پژوهش‌ها نشان دادند، نرم‌افزارهای خزنده‌نمایه‌ساز موتورهای کاوش وب در هنگام رویارویی با عناصر فراداده‌ای پیاده‌سازی شده در بستر زبان نشانه‌گذاری گسترش‌پذیر که در قالب پیشینه‌هایی مجزا، و یا جاسازی‌شده در اشیای محتوایی منتشر شده بودند، رویکرد حذف برچسب را اتخاذ می‌کنند. بدین معنی که نام‌های برچسب (عنصر) فاقد ارزش نمایه‌ای قلمداد شده، و به نمایه‌سازی ارزش‌های آنها بسنده می‌شود.

دلیل اتخاذ چنین رویکردی توسط موتورهای کاوش، ویژگی گسترش‌پذیر بودن قالب ایکس‌ام‌ال برای استفاده و تولید برچسب‌های موردنیاز و دلخواه تولیدکنندگان اشیای محتوایی است که طراحی الگوریتم‌های مناسب برای نمایه‌سازی و رتبه‌بندی اشیای محتوایی را برای طراحان موتورهای کاوش را دشوار می‌سازد. در پژوهش‌های یاد شده پیشینه‌های مبتنی بر قالب فراداده‌ای مارک به‌عنوان غنی‌ترین استاندارد فراداده‌ای با کارکرد توصیفی و مدیریتی نیز موردمطالعه قرار گرفت که دلیلی بر انعطاف‌پذیری قالب ایکس‌ام‌ال در امکان پیاده‌سازی تمامی استانداردهای فراداده‌ای، و حذف محدودیت‌های قالب پیشین (اچ‌تی‌ام‌ال) بود. پژوهش‌های این گروه، هر دو محیط انتشار اشیای محتوایی وب یعنی محیط‌های ایستا و پویا را مدنظر قراردادند، و توانایی هر یک از موتورهای کاوش موردمطالعه در نمایه‌سازی اشیای محتوایی دسترس‌پذیر در محیط‌های ایستا و پویا را مشخص نمودند. ویژگی‌هایی چون گسترش‌پذیری، ساختار سلسله مراتبی، و خودتوصیف بودن ایکس‌ام‌ال سطح بالایی از سازگاری با تعداد عناصر، ساختار، و بستر معناشناختی استانداردهای فراداده‌ای از یک‌سو، و قابلیت پردازش‌های بیشتر بر روی برچسب‌ها توسط ابزارهای بازیابی نظیر موتورهای کاوش وب از سوی دیگر را نشان می‌دهد.

نتایج پژوهش‌های حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب بیانگر آن است که استفاده از زبان نشانه‌گذاری گسترش‌پذیر، برخلاف زبان نشانه‌گذاری فرامتن، موجب بهبود فرایند میانکنش‌پذیری میان دو نظام درگیر در فرایند می‌گردد. برچسب‌های محدود اچ‌تی‌ام‌ال برای نشانه‌گذاری محتوای متنی، امکان پیاده‌سازی طرح‌های فراداده‌ای استاندارد با تعداد عناصر و ساختار معناشناختی غنی مانند قالب مارک را به وجود نمی‌آورند، و سطح نمایه‌پذیری عناصر طرح‌های فراداده‌ای استاندارد مبتنی بر این قالب را نیز به‌شدت کاهش می‌دهند. درحالی‌که ویژگی‌های زبان نشانه‌گذاری

گسترش‌پذیر محدودیت‌های بستر پیشین را مرتفع نموده، باعث افزایش دسترس‌پذیری به پیشینه‌های فراداده‌ای و اشیای محتوایی می‌شود.

اگرچه بستر نحوی جدید موجب بازیابی‌پذیری ارزش‌های عناصر مربوط به طرح‌های فراداده‌ای استاندارد می‌شود، اما انتخاب رویکرد حذف برچسب توسط موتورهای کاوش در هنگام نمایه‌سازی عناصر مبتنی بر ایکس‌ام‌ال، کارایی بازیابی را کاهش می‌دهد. به‌عبارت‌دیگر، کاربران نمی‌توانند درخواست‌های جستجو را به عناصری خاص محدود نموده، مانعیت نظام بازیابی را افزایش دهند. بازیابی حجم زیادی از اشیاء در برابر درخواست‌های ارائه‌شده به موتورهای کاوش بر شدت این مسئله می‌افزاید. به نظر می‌رسد، انجام پژوهش در خصوص بررسی نمایه‌سازی و پیدانمایی نام‌های (برچسب) عناصر توسط موتورهای کاوش وب با هدف بهبود فرایند بازیابی بتواند راهگشا باشد.

روش‌شناسی

چنان‌که پیشتر بیان شد، هدف اصلی پژوهش، تبیین تأثیر حذف خط اعلان زبان نشانه‌گذاری گسترش‌پذیر، و تغییر پسوند فایل برافزایش نمایه‌پذیری و پیدانمایی نام‌های برچسب عناصر فراداده‌ای هسته دوبلین، مارک ۲۱، و طرح فراداده‌ای توصیف شیء (MODS) در محیط موتورهای کاوش وب بود. به‌منظور نیل به هدف پژوهش حاضر که از لحاظ نوع یک پژوهش کاربردی به شمار می‌رود، چراکه بر حل مسئله تأثیر حذف خط اعلان ایکس‌ام‌ال از پیشینه‌های فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر، تغییر پسوند فایل پیشینه‌ها بر نمایه‌پذیری و پیدانمایی نام‌های برچسب عناصر فراداده‌ای، و امکان جستجوهای مبتنی بر عنصر برای بهبود بازیابی پیشینه‌های فراداده‌ای از طریق موتورهای کاوش وب در شرایط واقعی و عملی تأکید دارد، از روش‌های تجربی (از دو گروه گواه و آزمون بهره می‌برد)، و مبتنی بر فن‌آوری (در بستر فن‌آوری‌های نوین اطلاعاتی انجام‌شده و به مطالعه آنها می‌پردازد) استفاده شد.

در همین راستا، دو گروه سیصد پیشینه‌ای در قالب گروه‌های گواه و آزمون طراحی گردیدند. سیصد پیشینه گروه گواه دقیقاً بر پایه فرانماهای طرح‌های فراداده‌ای هسته دوبلین، مارک ۲۱، و طرح فراداده‌ای توصیف شیء (هر طرح فراداده‌ای صد پیشینه)، و سیصد پیشینه گروه آزمون (همانند گروه گواه برای هر طرح صد پیشینه)، مبتنی بر فرانماهای یاد شده اما بدون خط اعلان زبان نشانه‌گذاری گسترش‌پذیر و با پسوند فایل همخوان با طرح مبنای تولید، ایجاد شدند. اعتبار پیشینه‌های عضو گروه گواه بر پایه نرم‌افزارهایی خاص (Styles Studio و MARCXML MARC Validation Stylesheets) موردسنجش، و تأیید قرار گرفت. تلاش برای سازگاری کامل پیشینه‌های گروه گواه با فرانماهای طرح‌های فراداده‌ای

موردمطالعه بر پایه نتایج پژوهش‌های لوک و همکاران (Luk et al., 2000)، طاهری و همکاران (2009)، آقاعابدی (2010)، و طباطبایی امیری و همکاران (2011) بود، که نشان دادند رویکرد موتورهای کاوش برای نمایه‌سازی اشیای محتوایی مبتنی بر ایکس‌ام‌ال، رویکرد «حذف برچسب» است.

برای تولید پیشینه‌های فراداده‌ای گروه آزمون، تغییراتی خاص بر روی آنها اعمال شد. خط اعلان شیء مبتنی بر ایکس‌ام‌ال (<?xml version="1.0" encoding="UTF-8"?>) از ابتدای هر یک از پیشینه‌ها حذف شد، و پسوند فایل هر پیشینه نیز به نام خلاصه طرح فراداده‌ای مبنای تولید (dcxml.، marcxml.، یا mods.) تغییر یافت. لازم به ذکر است این تغییرات هیچ تأثیری بر قابلیت‌های بستر نحوی زبان نشانه‌گذاری گسترش‌پذیر ندارد، و پیشینه‌های گروه آزمون تمامی ویژگی‌های قبلی خود را حفظ می‌کردند. برای انتشار پیشینه‌های یاد شده بر روی وب، به هر یک از گروه‌های گواه و آزمون، وب‌سایتی مستقل با شرایطی کاملاً یکسان، و کنترل‌شده توسط پژوهشگر اختصاص یافت. هیچ پیوندی از وب‌سایت‌های دیگر به آن دو وب‌سایت انجام نشد، و هیچ پیوندی نیز از آن دو وب‌سایت به دیگر وب‌سایت‌ها تهیه نگردید.

سپس هر یک از وب‌سایت‌های گروه گواه (http://www.tagnamemeta-cont.ir) و گروه آزمون (http://www.tagnamemeta-ex2.ir) بر پایه بهترین روش‌های معرفی‌شده (نقشه‌های سایت مبتنی بر اچ‌تی‌ام‌ال، مبتنی بر متن، منبعِ منبع ، و مبتنی بر ایکس‌ام‌ال) به موتورهای کاوش گوگل و یاهو معرفی شدند. هر دو وب‌سایت از دی‌ماه ۱۳۹۰ تا دی‌ماه ۱۳۹۱ به مدت یکسال دسترس‌پذیر هستند. دلایل انتخاب طرح‌های فراداده‌ای هسته دوبلین، مارک ۲۱، و توصیف شیء، ویژگی ساختاری هر یک در قالب ایکس‌ام‌ال که ممکن است بر روی پردازش‌های آنها توسط ماشین، و به پیروی از آن نمایه‌سازی و پیدانمایی آنها تأثیرگذار باشد، کارکردپذیری هر یک، و کاربرد گسترده از آنها در نظام‌های اطلاعاتی بافت فرهنگی (کتابخانه‌ها، آرشیوها، و موزه‌ها) بودند.

بخش عمده پیشینه‌های طرح فراداده‌ای توصیف شیء از وب‌سایت کتابخانه کنگره (http://lccn.loc.gov)، و بخشی هم از طریق تبدیل پیشینه‌های فراداده‌ای مارک ۲۱ در بستر ایکس‌ام‌ال با استفاده از ابزار «MARCXML Toolkit» تهیه شدند. پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ نیز از وب‌سایت http://www.archive.org و از مجموعه اشیای محتوایی کتابخانه دیجیتالی کالیفرنیا، و از یک رده موضوعی(Authors) موردعلاقه پژوهشگر، با داشتن یک‌صد شیء محتوایی، و بدون تأثیر بر زمینه و نتایج انتخاب شدند. به‌منظور بررسی تأثیر حذف خط اعلان شیء ایکس‌ام‌ال و تغییر پسوند فایل به

نام طرح فرادادهای مبنای تولید برافزایش نمایه‌پذیری و پیدانمایی نام‌های برچسب عناصر فرادادهای در محیط موتورهای کاوش وب، می‌بایست مناسب‌ترین موتورهای کاوش وب در راستای تحقق هدف پژوهش انتخاب می‌شدند. بر پایه معتبرترین منابع موجود که با بهره‌گیری از معیارهایی چند به رتبه‌بندی موتورهای کاوش وب می‌پردازند، موتورهای کاوش گوگل و یاهو در بین موتورهای کاوش وب دارای بیشترین کیفیت خدمات، و بیشترین بسامد استفاده هستند (Alexa, 2012; Lewis, 2012; Campex, 2012; ComScore, 2012). ازاین‌رو در این پژوهش، دو موتور اشاره‌شده به‌عنوان نماینده موتورهای کاوش وب برگزیده شدند.

برای گردآوری داده‌ها از روش مشاهده طراحی‌شده استفاده شد. ابزار گردآوری داده‌ها، سیاهه وارسی (چک‌لیست) بود که با استفاده از نرم‌افزار آماری SPSS به‌عنوان ابزار ثبت مکانیکی توسط پژوهشگر طراحی شد. همچنین در مواردی از عکس‌های فوری ، و ذخیره آنها از طریق نرم‌افزار پینت، به‌عنوان ابزار الکترونیکی استفاده شد. روایی سیاهه وارسی بر پایه قضاوت پنج نفر از صاحب‌نظران باتجربه حوزه ذخیره و بازیابی اطلاعات مورد تأیید قرار گرفت. با توجه به استفاده از سیاهه وارسی به‌عنوان ابزار گردآوری داده‌ها، نیازی به سنجش پایایی نبود. فرایند گردآوری داده‌ها از اواسط بهمن‌ماه ۱۳۹۰ تا پایان اسفندماه ۱۳۹۰انجام شد.

تجزیه‌وتحلیل داده‌ها

برای آزمون فرضیه پژوهش که بر فراهم شدن امکان نمایه‌سازی و پیدانمایی نام‌های برچسب عناصر پیشینه‌های فرادادهای بدون سطر اعلان شیء ایکس‌ام‌ال، و با پسوند فایل همخوان با نام طرح فرادادهای مبنای تولید پیشینه‌ها مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر تأکید می‌نمود، متناسب با رویکرد تجربی پژوهش و نوع داده‌های به‌دست‌آمده، و نیز دوحالتی بودن متغیر پژوهش با توزیع زوجی و با یک احتمال مشخص (نمایه‌سازی و پیدانمایی نام‌های برچسب)از آزمون «مقایسه زوجی » استفاده شد.

جدول ۱. مقایسه زوجی بین نمایه‌پذیری و پیدانمایی نام‌های برچسب عناصر پیشینه‌های طرح‌های فراداده‌ای موردمطالعه عضو گروه آزمون و گروه گواه در محیط موتورهای کاوش گوگل و یاهو

گروه‌ها	موتورهای کاوش و طرح‌های فراداده‌ای موردمطالعه	آزمون =۱؛ گواه= ۲	رسته[1]	پیشینه	نسبت مشاهده‌شده[2]	نسبت آزمون[3]	معنی‌داری مجانبی (۲ دمی)[4]
گواه و آزمون	موتور کاوش گوگل – طرح فراداده‌ای هسته دوبلین	گروه ۱	۱	۱۰۰	۱,۰۰	.۵۰	.۰۰۰ (a)
		گروه ۲	۲	۱۰۰	.۰۰	.۵۰	
		مجموع		۲۰۰	۱,۰۰	۱,۰۰	
	موتور کاوش یاهو – طرح فراداده‌ای هسته دوبلین	گروه ۱	۱	۱۰۰	۱,۰۰	.۵۰	.۰۰۰ (a)
		گروه ۲	۲	۱۰۰	.۰۰	.۵۰	
		مجموع		۲۰۰	۱,۰۰	۱,۰۰	
	موتور کاوش گوگل- قالب فراداده‌ای مارک۲۱	گروه ۱	۱	۱۰۰	۱,۰۰	.۵۰	.۰۰۰ (a)
		گروه ۲	۲	۱۰۰	.۰۰	.۵۰	
		مجموع		۲۰۰	۱,۰۰	۱,۰۰	
	موتور کاوش یاهو – قالب فراداده‌ای مارک۲۱	گروه ۱	۱	۱۰۰	۱,۰۰	.۵۰	.۰۰۰ (a)
		گروه ۲	۲	۱۰۰	.۰۰	.۵۰	
		مجموع		۲۰۰	۱,۰۰	۱,۰۰	
	موتور کاوش گوگل –	گروه ۱	۱	۱۰۰	۱,۰۰	.۵۰	.۰۰۰ (a)

1. Category
2. Observed proportion
3. Test proportion
4. Asymptotic significance 2-tailed

	طرح فرادادهای توصیف شیء	گروه ۲	۲	۱۰۰	.۰۰	.۵۰	
		مجموع		۲۰۰	۱,۰۰	۱,۰۰	
	موتور کاوش یاهو – طرح فرادادهای توصیف شیء	گروه ۱	۱	۱۰۰	۱,۰۰	.۵۰	.۰۰۰ (a)
		گروه ۲	۲	۱۰۰	.۰۰	.۵۰	
		مجموع		۲۰۰	۱,۰۰	۱,۰۰	

همان‌طور که جدول۱ نشان می‌دهد، بر پایه نتیجه به‌دست‌آمده از آزمون آماری مقایسه زوجی، تفاوت معناداری بین نمایه‌سازی و پیدانمایی نام‌های برچسب عناصر پیشینه‌های مبتنی بر طرح‌های فرادادهای موردمطالعه در گروه آزمون با گروه گواه وجود دارد. به‌نحوی‌که در تمامی آن پیشینه‌ها مشاهده‌شده، و ازلحاظ آماری معنادار است. بنابراین فرضیه پژوهش تأیید می‌شود. استفاده از رویکرد استفاده‌شده در طراحی پیشینه‌های گروه آزمون یعنی حذف خط اعلان شیء مبتنی بر ایکس‌ام‌ال و تغییر پسوند فایل مربوط به پیشینه‌های فرادادهای، موجب نمایه‌سازی و پیدانمایی نام‌های برچسب آنها توسط موتورهای کاوش گوگل و یاهو به‌عنوان نمایندگان موتورهای کاوش وب شده است.

نتیجه

نمایه‌سازی نام‌های برچسب عناصر فرادادهای توسط موتورهای کاوش وب، و نمایش آنها در نتایج جستجو، امکان جستجوی پیشینه‌های فرادادهای بر پایه عناصر آنها را امکان‌پذیر نموده، موجب بهبود دسترسی به اشیای محتوایی وب می‌شود. بدین‌صورت که دقت و مانعیت فرایند بازیابی را افزایش داده، اشیای مرتبط‌تری را در برابر درخواست‌های ارائه‌شده به نظام بازیابی ارائه می‌نماید. این امکان در محیط موتورهای کاوش وب که پیشینه‌های فراوانی را در برابر درخواست‌های کاربران ارائه می‌دهند، ارزش‌افزوده به شمار می‌آید. ازسوی‌دیگر، امکان جستجوهای مبتنی بر عناصر فرادادهای در موتورهای کاوش، موجب افزایش دسترس‌پذیری اشیای محتوایی سازمان‌یافته موجود در پایگاه‌های اطلاعاتی تمام متن و پایگاه‌های کتابشناختی (فهرست‌های رایانه‌ای) بافت اطلاعات فرهنگی (کتابخانه‌ای، آرشیوی، و موزه‌ای)، آن هم در محیط رابط کاربری واحد (و موردعلاقه کاربران) خواهد شد. بر این اساس، موتورهای کاوش همانند دروازه‌های اطلاعاتی برای پایگاه‌ها و شبکه‌های اطلاعاتی عمل خواهند نمود، و

کاربران را از مراجعه جداگانه به هر یک از پایگاه‌های اطلاعاتی بی‌نیاز خواهند ساخت. به‌عبارت‌دیگر، امکان جستجوی یکپارچه در پایگاه‌ها و شبکه‌های اطلاعاتی در محیط پراستفاده‌ترین ابزار کاوش وب، بدون محدودیت به پلت‌فرم نرم‌افزاری خاص فراهم خواهد شد.

تغییر بستر نحوی پیاده‌سازی پیشینه‌های فراداده‌ای در پژوهش‌های پیشینی که زبان ایکس‌ام‌ال را به‌عنوان بستر نحوی موردمطالعه قرار داده بودند، افزایش سطح نمایه‌پذیری و پیدانمایی عناصر فراداده‌ای را در پی داشت. موتورهای کاوش وب ارزش‌های عناصر پیشینه‌های فراداده‌ای موردبررسی را به‌صورت کامل نمایه‌سازی کرده، و به‌وسیله نظام جستجوی خود قابل بازیابی ساختند. این مهم هم در مورد پیشینه‌های منتشر شده در محیط‌های ایستا و هم پیشینه‌های منتشر شده در محیط‌های پویا رخ داد. بااین‌وجود، ساختار معمولی پیشینه‌ها، بدون رعایت هیچ ویژگیِ خاصی، تنها شرایط را برای نمایه‌پذیری و پیدانمایی ارزش‌های عناصر مهیا نموده، و موتورهای کاوش با نام‌های برچسب عناصر به‌عنوان اصطلاحات نمایه‌ای برخورد نکرده، و رویکرد حذف برچسب را برگزیدند (Taheri et al., 2009; Agha Abedi, 2010; Tabatabaie Amiri et al., 2012).

این در حالی است که یافته‌های پژوهش حاضر نشان می‌دهد، راهبرد استفاده‌شده برای تولید پیشینه‌های فراداده‌ای مبتنی بر ایکس‌ام‌ال، یعنی حذف خط اعلان ایکس‌ام‌ال و تغییر پسوند فایل به نام طرح فراداده‌ای مبنای تولید موجب می‌شود، موتورهای کاوش گوگل و یاهو در برخورد با پیشینه‌های فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر موردمطالعه، رویکرد حذف برچسب خود را تغییر دهند. به‌عبارت‌دیگر، نام‌های برچسب عناصر را به‌مثابه اصطلاحات نمایه‌ای نمایه‌سازی نموده، و به آنها اجازه پیدانمایی در برابر درخواست‌ها را دهند. این شیوه برخورد با نام‌های برچسب عناصر، بهترین رویکردی است که در منابع مختلف برای نمایه‌سازی اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر توسط موتورهای کاوش بیان شده است (Luk et al., 2000; Qin, 2000; Luk et al., 2002; Gigee, 2006; Gill, 2008; Taheri and Hariri, 2012).

البته بهترین قالبی که موتورهای کاوش می‌توانند رویکرد برخورد با نام‌های برچسب اشیای محتوایی پیاده‌سازی شده در آن را به‌عنوان اصطلاحات نمایه‌ای انتخاب کنند نیز ایکس‌ام‌ال است. ویژگی گسترش‌پذیری این قالب نسبت به دیگر قالب‌ها و زبان‌های نشانه‌گذاری چنین مزیتی را فراهم نموده است. این در حالی است که قابلیت نمایه‌پذیری و پیدانمایی تمامی نام‌های برچسب عناصر تاکنون برای هیچ‌یک از بسترهای نحوی یا قالب‌های ذخیره‌سازی، انتقال، و نمایش داده‌ها در موتورهای کاوش وب وجود نداشته است. عدم نمایه‌پذیری و پیدانمایی نام‌های برچسب عناصر در مورد پیشینه‌های فراداده‌ای

موردمطالعه در پژوهش‌هایی که زبان نشانه‌گذاری فرامتن را به‌عنوان بستر پیاده‌سازی برگزیده بودند نیز صدق می‌کرد. یافته‌های این پژوهش‌ها بیانگر ضعف کارایی زبان نشانه‌گذاری فرامتن به‌عنوان بستر نحوی برای تولید پیشینه‌هایی با قابلیت نمایه‌سازی و پیدانمایی توسط موتورهای کاوش وب بود. محدودیت برچسب‌های این زبان که مانع پیاده‌سازی اغلب طرح‌های فراداده‌ای استاندارد می‌شد از یک‌سو، و واکنش مثبت موتورهای کاوش فقط به برخی فرابرچسب‌های این زبان، آن هم صرفاً ارزش‌های آنها و نه نام‌های برچسب، و نادیده گرفتن عناصر مبتنی بر طرح‌های استاندارد در هنگام فرایند نمایه‌سازی از سوی دیگر، کاهش میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش، و درنتیجه دسترسی به اشیای محتوایی را در پی داشت (Turner and Brackbill, 1998; Sokvitne, 2000; Henshaw and Valauskas, 2001; Quevedo-Torrero, 2004; Zhang and Dimitroff; 2004; Zhang and Dimitroff, 2005a; Zhang and Dimitroff, 2005b; Safari, 2005; Mohamed, 2006; Sharif, 2007).

انتخاب زبان نشانه‌گذاری گسترش‌پذیر به‌جای زبان نشانه‌گذاری فرامتن به‌عنوان بستر نحوی توسط طراحان استانداردهای فراداده‌ای، به‌منظور حذف محدودیت‌های بستر پیشین و استفاده از قابلیت‌های منحصربه‌فرد بستر جدید بود. نکته مهم در خصوص راهبرد به‌کاررفته در این پژوهش، حفظ قابلیت‌های قالب ایکس‌ام‌ال است. همان‌طور که بسیاری از منابع مرجع توصیف و آموزش زبان نشانه‌گذاری گسترش‌پذیر بیان می‌دارند، سطر اعلان شیء ایکس‌ام‌ال، سطری اختیاری است و صرفاً برای نشان دادن نسخه زبان ایکس‌ام‌ال، و مجموعه نویسه‌های مورداستفاده است، و تأثیری بر ویژگی‌های محتوایی شیء مرتبط ندارد. مرورگرهای وب نیز اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر بدون سطر اعلان را به‌طور کامل شناسایی و پردازش می‌کنند (Harold, 2012; Microsoft Developer Network, 2012; Validome, 2012; JCommerce Developer Network, 2012; Tizag, 2012).

تغییر پسوند فایل مربوط به پیشینه مبتنی بر ایکس‌ام‌ال نیز بر قابلیت‌های محتوایی شیء تأثیر نخواهد داشت. تغییر پسوند فایل، تغییر ظاهری بر روی فایل است و به محتوای آن، و ساختار و قابلیت‌های مورداستفاده آسیبی نمی‌رساند. آنچه در زبان‌های نشانه‌گذاری حائز اهمیت است، حفظ نشانه‌ها، ساختار، و روابط معناشناختی میان عناصر آنهاست. پیشینه‌های مبتنی بر ایکس‌ام‌ال با داشتن پسوند فایلی متفاوت نیز ویژگی‌های قبلی را خواهند داشت، و استفاده از قابلیت‌های پردازشی فناوری‌های مرتبط با ایکس‌ام‌ال برای آنها فراهم است. با این ویژگی مثبت که انتخاب نام خود طرح فراداده‌ای در بستر زبان نشانه‌گذاری گسترش‌پذیر، یعنی marcxml.، dcxml.، وmods.، شناسایی، دسته‌بندی، و درک آنها را توسط ماشین و انسان افزایش می‌دهد. حتی در پرطرفدارترین موتور کاوش وب (گوگل) نیز امکان محدود نمودن قالب

اشیای بازیابی شده به پسوند فایل‌های جدید وجود دارد. ازآنجاکه راهبرد به‌کاررفته در پژوهش حاضر برای تولید پیشینه‌های فرادادهای با نام‌های برچسب نمایه‌پذیر و قابل پیدانمایی که موجب بهبود بازیابی اشیای محتوایی در محیط موتورهای کاوش وب می‌شود، تأثیر صددرصدی داشت، الگوهایی چند به مخاطبان پژوهش، یعنی تولیدکنندگان پیشینه‌های فرادادهای و طراحان موتورهای کاوش وب، به‌منظور بهره‌مندی از مزایای آن‌ها در قالب پیشینه‌های فرادادهای در ذیل ارائه می‌گردند. دو الگوی نخست برای تولیدکنندگان پیشینه‌های فرادادهای، و دو الگوی دیگر برای طراحان موتورهای کاوش وب می‌باشند. لازم به ذکر است پیشینه‌های الگوی مبتنی بر قالب فرادادهای مارک ۲۱ و طرح فرادادهای توصیف شیء که در اصل پژوهش ارائه شده‌اند نیز شبیه الگوی نخست هستند.

الگوی نخست. نمونه‌ای از یک پیشینه مبتنی بر هسته دوبلین دارای ارزش‌های عناصر به زبان فارسی بدون سطر اعلان ایکس‌ام‌ال و با پسوند فایل «.dcxml»

چنان‌که قبلاً گفته شد، مرورگرهای وب (ازجمله اینترنت اکسپلورر و فایرفاکس) قادر به نمایش اشیای مبتنی بر ایکس‌ام‌ال به شکل طبیعی هستند، و پیشینه‌های تولیدشده ازلحاظ نمایشی تفاوتی با دیگر پیشینه‌ها ندارند. همان‌طور که در الگوی بالا مشاهده می‌شود، پیشینه نمونه فاقد سطر اعلان ایکس‌ام‌ال است، و اولین عنصری که در آن نمایان شده، عنصر ریشه طرح فرادادهای هسته دوبلین است. این پیشینه نمونه، کاملاً ساختار و ویژگی‌های اشیای مبتنی بر ایکس‌ام‌ال را حفظ می‌کند. پسوند فایلی که پیشینه در آن ذخیره‌شده، همخوان با نام طرح فرادادهای است (Extenstion=.dcxml). این پسوند هیچ تأثیری بر روی محتوای فایل نخواهد داشت. محتوای این پیشینه‌ها در هر دو مرورگر پرکاربرد وب (اینترنت اکسپلورر و فایرفاکس) به‌صورت یکسان به نمایش درمی‌آیند، و شکل ظاهری آنها با پیشینه‌های مبتنی بر زبان نشانه‌گذاری گسترش‌پذیری که بر پایه ساختار معمولی پیشینه‌ها تولیدشده‌اند، یکسان است .

با توجه به کاربردپذیری کامل الگوی نخست، و نیز استفاده محدود موتورهای کاوش وب از نظام‌های معنایی(صرفاً شبکه‌های معنایی یا اصطلاحنامه‌های زبانی و نه تخصصی) و عدم استفاده از فهرست‌های مستند، طراحی الگویی مناسب و کاربردی برای افزایش جامعیت نظام بازیابی موتورهای کاوش امکان‌پذیر است. در این الگو، پیشینه‌های مستند (نام‌ها، موضوعی، و ...) بر پایه رویکرد «پیشینه‌های مجموعه » در پیشینه کتابشناختی مرتبط جاسازی‌شده، و ارزش‌ها و نام‌های برچسب عناصر آنها نیز توسط موتورهای کاوش نمایه‌سازی شده، و به سطح پیدانمایی می‌رسد. بدین ترتیب قابلیت جستجوی دیگر

شکل‌های نام‌ها و اصطلاحات (شکل‌های پذیرفته‌نشده) نیز برای کاربران میسر می‌شود. این ویژگی برای حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب ارزش‌افزوده در پی خواهد داشت.

الگوی دوم. نمونه‌ای از یک پیشینه مجموعه مبتنی بر هسته دوبلین، ترکیبی از پیشینه کتابشناختی و پیشینه مستند مرتبط با آن، بدون سطر اعلان ایکس‌ام‌ال و با پسوند فایل ".dcxml"

در الگوی دوم، پیشینه مستند موضوعی مبتنی بر طرح فراداده‌ای توصیف مستند (MADS) که از پایگاه مستند موضوعی کتابخانه کنگره آمریکا استخراج‌شده است، و پیشینه کتابشناختی تولیدشده بر پایه طرح فراداده‌ای هسته دوبلین مرتبط با آن در یک پیشینه مجموعه جاگرفته‌اند (بدیهی است امکان طراحی و ایجاد پیشینه‌های مجموعه مبتنی بر دیگر طرح‌های فراداده‌ای کتابشناختی نیز وجود دارد). این الگو نیز هم از لحاظ فنی و هم از لحاظ رویکردهای حوزه فراداده معتبر بوده، و از کارایی کامل برخوردار است. لازم به ذکر است برای جاسازی پیشینه‌های مستند موضوعی در پیشینه‌های مجموعه، می‌بایست اصطلاحاتی که با اصطلاح پذیرفته‌شده رابطه سلسله مراتبی یا وابسته دارند از پیشینه حذف شوند، و فقط اصطلاحات هم‌ارز باقی بمانند. در غیر این صورت، جستجوی اصطلاحات مرتبط، موجب بازیابی پیشینه‌هایی نامرتبط، یا به‌عبارت‌دیگر، ریزش کاذب خواهد شد.

در ذیل الگویی برای تغییر رویکرد موتورهای کاوش (حذف برچسب) به نمایه‌سازی هوشمند نام‌های برچسب عناصر پیشینه‌های فراداده‌ای، و ارائه آنها در نتایج جستجو ارائه‌شده است.

الگوی سوم. عنصر ریشه طرح فراداده‌ای هسته دوبلین به‌عنوان معیار تشخیص پیشینه فراداده‌ای برای موتورهای کاوش

الگوی ارائه‌شده، برچسب‌های آغازین و پایانی عنصر ریشه در طرح فراداده‌ای هسته دوبلین است. موتور کاوش در هنگام برخورد با پیشینه‌های دارای این برچسب‌ها، برچسب‌ها و ارزش‌های میان آنها را عناصر پیشینه‌های فراداده‌ای محسوب نموده، نام‌های برچسب عناصر مرتبط را نمایه‌سازی، و در نتایج جستجو نمایان می‌سازد. بدیهی است چنین حالتی هم برای پیشینه‌های فراداده‌ای پیاده‌سازی شده در بستر نحوی زبان نشانه‌گذاری گسترش‌پذیر با ساختار معمولی، و هم برای پیشینه‌های تولیدشده بر پایه راهبرد به‌کاررفته در این پژوهش قابل اجراست. همچنین رویکرد استفاده از فراداده، پیوند با شیء محتوایی یا جاسازی آن در شیء (در برچسبی مانند <meta>) نیز تفاوتی ایجاد نخواهد کرد. چنین عناصری در دو طرح فراداده‌ای دیگر موردمطالعه در این پژوهش، و نیز طرح‌های فراداده‌ای نظیر «استاندارد انتقال و کدگذاری فراداده (METS)»، «فراداده برای مدیریت تصاویر ثابت (MIX)»، و نظیر آنها نیز وجود دارد، و

به‌عنوان معیار برای موتورهای کاوش کاربردپذیر است. محدود بودن تعداد استانداردهای فراداده‌ای، کاربردپذیری الگو برای موتورهای کاوش را بسیار تسهیل می‌نماید.

نمایه‌سازی نام‌های برچسب عناصر فراداده‌ای و ارائه آنها در نتایج جستجو برای موتورهای کاوش از جنبه‌های مختلفی ارزش‌افزوده محسوب می‌شود. نخست، کیفیت ارزش‌های ذخیره‌شده در پیشینه‌های فراداده‌ای به‌صورت ساختاریافته، غنای محتوای پایگاه داده‌های موتورهای کاوش را بهبود می‌بخشد. سپس، کاربردپذیری موتورهای کاوش را نزد پژوهشگران به‌عنوان کاربران سطح عالی وب افزایش می‌دهد. دیگر اینکه، موتورهای کاوش را به دروازه‌های اطلاعاتی با تضمین اطلاعات کیفی تبدیل می‌کند.

پیشنهاد دیگری که می‌توان بر پایه یافته‌های مربوط به گروه آزمون، و پژوهش‌های طاهری و همکاران (2009)، آقاعابدی (2010)، و طباطبایی امیری و همکاران (2011) به‌عنوان الگو به موتورهای کاوش ارائه نمود، استفاده از فرانمای استانداردهای فراداده‌ای است.

بر پایه این الگو، همانند رویکرد پایگاه‌های اطلاعاتی مبتنی بر ایکس‌ام‌ال (به‌صورت قالب محلی یا قابلیت پشتیبانی)، فرانمای ایکس‌ام‌ال طرح فراداده‌ای به پایگاه موتور کاوش معرفی‌شده، نرم‌افزار روبات با مشاهده پیشینه‌هایی مبتنی بر این فرانما، آنها را پیشینه‌های فراداده‌ای قلمداد نموده، نام‌های برچسب آنها را نمایه‌سازی می‌کند. نرم‌افزار پردازشگر درخواست نیز، راهبردهای دارای نام برچسب را معتبر ارزیابی می‌نمایند. فرانماهای یکی از طرح‌های فراداده‌ای موردمطالعه به‌عنوان الگوهای چهارم مقاله پیشنهاد می‌گردند:

الگوی چهارم. فرانمای سطح ویژه طرح فراداده‌ای هسته دوبلین به‌عنوان ابزاری برای پردازش پیشینه‌های فراداده‌ای توسط موتورهای کاوش

انتخاب سه طرح فراداده‌ای معتبر با ساختارهای متفاوت و تأثیرگذار بر فرایند پردازش، تعداد پیشینه‌های موردمطالعه، و نتیجه آزمون انجام‌شده بر روی فرضیه، امکان تعمیم‌یافته‌های پژوهش را فراهم می‌نماید. بنابراین الگوهای ارائه‌شده در مقاله می‌تواند برای طراحی و تولید پیشینه‌های مبتنی بر دیگر استانداردهای فراداده‌ای بافت اطلاعات فرهنگی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر مورداستفاده قرار گیرد. بدین ترتیب کتابخانه‌ها، آرشیو، موزه‌ها، و دیگر سازمان‌هایی که در ارائه خدمات اطلاعاتی فعالیت می‌کنند قادر خواهند بود، پیشینه‌های فراداده‌ای اصلی خود و یا نسخه‌ای دیگر از آن پیشینه‌ها را بر پایه الگوهای ارائه‌شده تولید، و از مزیت‌های آن بهره‌مند گردند.

در مقایسه با دیگر ابزارهای دسترسی به محتوای وب، موتورهای کاوش به دلیل محبوبیتی که نزد کاربران وب دارند، و نظام‌های فراداده‌ای به دلیل قابلیت‌هایی که در توصیف و دسترسی دقیق به اشیای محتوایی دارا هستند، از اهمیت و جایگاه بالاتری برخوردارند. تعامل بیشتر این دو نظام، بر پایه نگرش ترکیب ابزارهای ذخیره و بازیابی محتوای وب (ناشی از عدم پاسخگویی هر یک از ابزارها به‌تنهایی)، دسترس‌پذیری بیشتر اشیای محتوایی وب را باعث می‌شود. در همین راستا، الگوهای پیشنهادی پژوهش بر گسترش میانکنش‌پذیری میان نظام‌های فراداده‌ای و موتورهای کاوش وب تأکید دارند. این الگوها متناسب با هر دو گروه درگیر در این تعامل یعنی تولیدکنندگان پیشینه‌های فراداده‌ای و طراحان موتورهای کاوش طراحی شدند، تا راه‌حل‌هایی جامع را ترسیم کنند. انجام پژوهش‌های بیشتر در حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش، ترکیب بیشتر دو نظام را به‌عنوان ابزاری جدید و کارآمد برای دسترسی به محتوای وب به ارمغان خواهد آورد. بررسی کارآمدی فراپیشینه‌های طراحی‌شده بر مبنای الگوی مفهومی ملزومات کارکردی پیشینه‌های کتابشناختی با خصایصی خاص در بستر زبان نشانه‌گذاری گسترش‌پذیر برای بهبود بازیابی معنادار سطوح مختلف اشیای محتوایی (اثر، بیان، بازنمون، و مورد) و اشیای مرتبط با یکدیگر پیشنهاد می‌شود.

References

Alexa: The Web Information Company (2011). Global Top 500. Retrieved 14 Dec. 2011 from: http://www.alexa.com/site/ds/top_sites?ts_mode=global&lang=none

Aqa Abedi, E. (2012). The effect of syntax on the indexing & ranking of metadata records by the web search engine: a comparative study on MARCXML and DCXML metadata records, unpublished Master’s thesis, Science and Research Branch, Islamic Azad University, Tehran. (In Persian)

Bogaards, Peter J. (2004). Metadata and XML: Improving the Findability of Information. Retrieved 7 Feb. 2012 from: http://www.bogieland.com/taxi/metadata_and_xml.pdf

Bray, Tim; Paoli, Jean; Sperberg-McQueen, C. M.; Maler, Eve; Yergeau, François (2008). Extensible Markup Language (XML) 1.0 (Fifth Edition): W3C Recommendation 26 November 2008. Retrievd 7 Feb. 2012 from: http://www.w3.org/TR/xml/

Campex (2012). Top Search Engines. Retrieved 7 Feb. 2012 from: http://capmex.biz/resources/top-search-engines

ComScore (2012). comScore Releases May 2012 U.S. Search Engine Rankings. Retrieved 7 Feb. 2012 from: http://www.comscore.com/Press_Events/Press_Releases/2012/6/comScore_Releases_May_2012_U.S._Search_Engine_Rankings

Farance, Frank; Gillman, Dan (2007). Not as hard as it sounds: Using XML in Metadata-enabled infrastructure. Retrieved 5 Nov. 2011 from:http://jtc1sc32.org/doc/N1851-1900/32N1896-ISO_Focus-metadata.pdf

Gigee, Grant (2006). MARC and MARCXML.Retrieved 5 Nov. 2011 from:http://threegee.files.wordpress.com/2006/05/marcxml.pdf

Gill, Toney (2008).Metadata and the Web: Introduction to Metadata. Retrieved 5 Nov. 2011 from:http://www.getty.edu/research/publications/electronic_publications/intrometadata/metadata.pdf

Hariri, N.; Taheri, S.M. and Fattahi, S.R. (2014). Interoperability between metadata systems and web search engines: Current Development and Trends, Library and Information Research Journal. 3 (2). (In Persian)

Harold, Elliotte Rusty (2012). Effective XML. Retrieved 5 Dec. 2011 from:http://www.cafeconleche.org/books/effectivexml/chapters/01.html

Henshaw, Robin; Valauskas, Edward J (2001).Metadata as a Catalyst: Experiments with Metadata and Search Engines in the Internet Journal, First Monday. Retrieved 14 Dec. 2011 fromwww.librijournal.org/pdf/1999-3pp125-131.pdf

Hirwade, Mangala Anil (2011). A study of metadata standards. Library Hi Tech News, Vol. 28 Iss: 7 pp. 18- 25

JCommerce Developer Network (2012).Version declaration.Retrieved 14 Dec. 2011 from:http://www.javacommerce.com/displaypage.jsp?name=version.sql&id=18238

Lewis, Edward (2008). Top Ten Search Engines. Retrieved 14 Dec. 2011 from http://www.seoconsultants.com/search-engines/

Luk, Robert; Chan, Alvin; Dillon, Tharam; Leong, H. V. (2000).A Survey of Search Engines for XML Documents. Retrieved 14 Dec. 2011 from: http://www.haifa.il.ibm.com/sigir00-xml/final-papers/Luk/XMLSUR.htm

Luk, Robert; Leong, H. V.; Dillon, Tharam S.; Chan, Alvin T. S.; Croft, W. Bruce; Allan, James (2002). A Surveyin Indexing and Searching XML Documetns.Retrieved 14 Dec. 2011 from: http://onlinelibrary.wiley.com/doi/10.1002/asi.10056/full

Mascaro, Michelle J (2004). The Value of Flexibility in Metadata Schemas. Retrieved 14 Dec. 2011 from: http://etd.ils.unc.edu:8080/dspace/bitstream/1901/57/1/michellemascaro.pdf

Microsoft Developer Network (MSDN) (2012). XML Declaration [XML Standards]. Retrieved 14 Feb. 2012 from: http://msdn.microsoft.com/en-us/library/ms256048.aspx.

Mohamed, KhaledA.f. (2006). The impact of metadata in web resources discovering.Online Information Review. 30 (2): 155-167

Qin, Jian (2000). Representation and Organization of Information in the Web Space: From MARC to XML. Retrieved 12 Dec. 2011 from: http://inform.nu/Articles/Vol3/v3n2p83-88.pdf

Quevedo-Torrero, Jesus Ubaldo (2004). IMPROVING WEB RETRIEVAL BY MINING THE HTML TAGS FOR KEYWORDS AND EXPLORING THE HYPERLINK STRUCTURES WEB PAGES.Ph. D. Dissertation, Department of Computer Science, University of Houston. Retrieved 12 Dec. 2011 from: http://wwwlib.umi.com/dissertations/fullcit/3156028

Safari, Mehdi (2005). Search Engine and Resource Discovery on the Web: Is Dublin Core an Impact Factor. Retrieved 12 Dec. 2011 from: www.webology.ir/2005/v2n2/a13.html

Search Engine Watch (2007). How Search Engines Work. Retrieved 12 Dec. 2011 from: http://searchenginewatch.com/article/2065173/How-Search-Engines-Work

Sharif, A. (2007). Study the effectiveness of metadata elements on web page visibility in publicsearch engines, available at: http://eprints.rclis.org/handle/10760/9171#.UHPNcVG94hA (accessed December 7, 2012). (In Persian)

Sokvine, Lloyd (2000). An Evaluation of the Effectiveness of Current Dublin Core Metadata for Retrieval. Retrieved 12 Dec. 2011 from: www.vala.org.au/vala2000/2000pdf/Sokvitne.PDF

Tabatabai Amiri F, Taheri S. M., Farajpahlou A, Osareh F, Moarrefzadeh A. (2012). Web Search Engines and Indexing and Ranking the Content Object Including Metadata Elements Available at the Dynamic Information Environments, Journal of Information Processing and Management. 27 (4). (In Persian)

Taheri, S. M.; Hariri, Nadjla (2012).A Comparative Study on the Indexing and Ranking of the Content Objects Including the MARCXML and Dublin Core's Metadata Elements by General Search Engines. Electronic Library.Vol 30, issue 4

Taheri, S.M. and Hariri, N. (2012). A comparative study on the indexing and ranking of the content objects including the MARCXML and Dublin Core’s metadata elements by general search engines, Library and Information Quarterly. No.: 48. (In Persian)

Tizag (2012).XML Tutorial. Retrieved 12 Dec. 2011 from: http://www.tizag.com/xmlTutorial/xmlprolog.php

Turner, Thomas P.; Brackbill, Lise (1998).Rising to the Top: Evaluating the Use of the HTML META Tag To Improve Retrieval of World Wide Web Documents through Internet Search Engines Retrieved 12 Dec. 2011 from: http://cat.inist.fr/?aModele=afficheN&cpsidt=1748620

Validome (2012).Error in the XML-Declaration.Retrieved 12 Dec. 2011 fromhttp://www.validome.org/lang/en/errors/XML-DECLARATION

Zhang, Jin; Dimitroff, Alexandra (2004).Internet search engine's response to metadata Dublin Core implementation. Retrieved 12 Dec. 2011 from http://portal.acm.org/citation.cfm?id=1142111

Zhang, Jin; Dimitroff, Alexandra (2005a).The impact of metadata implementation on Webpage visibility in search engine result (Part II). Retrieved 12 Dec. 2011 from: http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VC8-4BHCBX4-2&_user=10&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=a853d410a866732d3f8ab5dd3217d412

Zhang, Jin; Dimitroff, Alexandra (2005b). The impact of Webp age content characteristics on webpage visibility in search engine result (Part I). Retrieved 12 Dec. 2011 from: http://web.simmons.edu/~braun/467/part_1.pdf